# Lab #3

Cindy J. Pang Lab 1B, MWF 12:00-12:50pm Week 2 – August 12, 2024

## 1. Power Transformations and Assessing Normality

## Normality



#### Standard normal distribution

Usually, normally distributed data is the most ideal, It is well-understood and has many benefits:

- Easy to understand: mean as a central tendency and sd as a variance parameter
- Can be directly compared with other data
- Meet the assumptions of many statistical analysis.

But in many cases, our data is far from a normal distribution...

....

No. of standard deviations from the mean

### **Power Transformation**

Data transformation can transform non-normal data to approximately normal distribution. Common transformations include the **logarithm, square root, and square** 

### Log-transformation Example:

- Log Transformation is usually used when the data is **positively skewed**, where data in the long tail is far away from the center
- Applying a logarithmic transformation compresses the range by pulling in the extreme values more than the lower values, which makes it more symmetric.



### **Caution: Zero Values in Power Transformation**

=LN(D36)													
с	D	D E		G									
UARE	ORIGINAL	SQUARE ROOT	LOGS	RECIPROCAL ROOT	RI								
324	18	4.242640687	2.890372	0.23570226	0								
289	17	4.123105626	2.833213	0.242535625	0								
196	14	3.741657387	2.639057	0.267261242	0								
441	21	4.582575695	3.044522	0.21821789	0								
441	21	4.582575695	3.044522	0.21821789	0								
441	21	4.582575695	3.044522	0.21821789	0								
0	0	) 0	=LN(D36)	#DIV/0!									
196	14	3.741657387	2.639057	0.267261242	0								
841	29	5 385164807	3 367296	0 185695338	n								

- Many power transformation, such as log, is not applicable if the data contains 0's!
- It is important to check your original and transformed data to make sure there is no error produced

Log(0) = UNDEFINED 1/0 = UNDEFINED

CUBE	SQUARE	ORIGINAL	SQUARE ROOT	LOGS	RECIPROCAL ROOT	RECIPROCAL	RECIPROCAL SQUARE	RECIPROCAL C	UBE
5832	324	18	4.242640687	2.890372	0.23570226	0.055555556	0.00308642	0.000171468	
4913	289	17	4.123105626	2.833213	0.242535625	0.058823529	0.003460208	0.000203542	
2744	196	14	3.741657387	2.639057	0.267261242	0.071428571	0.005102041	0.000364431	
9261	441	21	4.582575695	3.044522	0.21821789	0.047619048	0.002267574	0.00010798	
9261	441	21	4.582575695	3.044522	0.21821789	0.047619048	0.002267574	0.00010798	
9261	441	21	4.582575695	3.044522	0.21821789	0.047619048	0.002267574	0.00010798	
0	0	0	Ø	#NUM!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	
2744	196	14	3.741657387	2.639057	0.267261242	0.071428571	0.005102041	0.000364431	
24389	841	29	5.385164807	3.367296	0.185695338	0.034482759	0.001189061	4.10021E-05	
24389	841	29	5.385164807	3.367296	0.185695338	0.034482759	0.001189061	4.10021E-05	- ··· ·

### Solutions for error messages in transformed data

 One way to omit the error: IF() combined with ISNUMBER() inside summary functions =SUM(IF(ISNUMBER(A1:A10), A1:A10))

### =PERCENTILE(IF(ISNUMBER(A1:A10), A1:A10), 0.25)

-In this case, IF() evaluates whether the value is a valid number for each value one by one, return True or False

-If True, this data is included, if false, it is not. The error message "#NUM!" is not a valid number, thus it will not be included in the summary functions

• For average(), you can also use **AVERAGEIFS()** instead:

=AVERAGEIFS(A1:A10, "<>#DIV/0!")

## Normality Test - Hypothesis Testing

The mechanism of normality test is beyond the scope, this diagram is just a simple illustration of a general hypothesis testing, you will learn more in the later lectures.

#### IMPORTANT: HOW TO INTERPRET P-VALUE



Slides Courtesy of Qingyuan Liu

Note: p-value is not a golden standard, the test result could be misleading

# 2. Union and Intersection

<u>Recall:</u> Axioms of Probability

- 1.
- 2.

### Union (U) vs. Intersection ( $\cap$ )





### Complements

- Denoted like E<sup>c</sup> or A<sup>c</sup> with the "c" subscript. When you see the "c" you can read it like "not E" or "not A"
- Other ways you may see this notation:
  - They all mean the same thing (not E)
- Examples:
  - $\circ A^c$

 $\circ \quad (A \cup B)^c$